

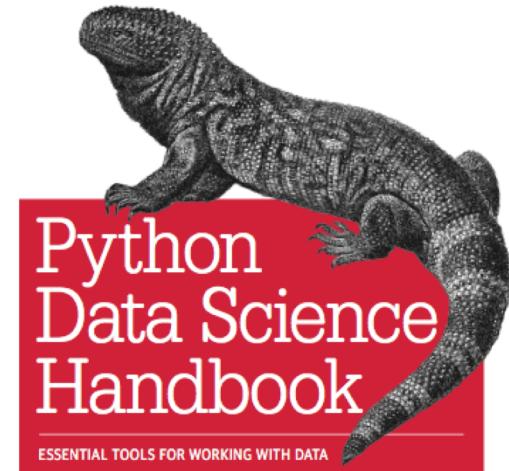
Programming for Data Science

CSC310



Course Details

- [Dr Lutz Hamel](#)
- Best way to get in touch - email:
 - lutzhamel@uri.edu
- Everything is online
 - Gradebook (BrightSpace)
 - Course Website
 - homepage.cs.uri.edu/faculty/hamel/courses/home/csc310/
 - Shared Google Drive/Github Repo
 - [Google Drive Link](#)
 - github.com/lutzhamel/ds
 - Lecture Notes
 - Example Code
- Python!
 - Python3
 - Google Colab Notebooks
 - colab.research.google.com/notebooks/intro.ipynb



powered by
 jupyter

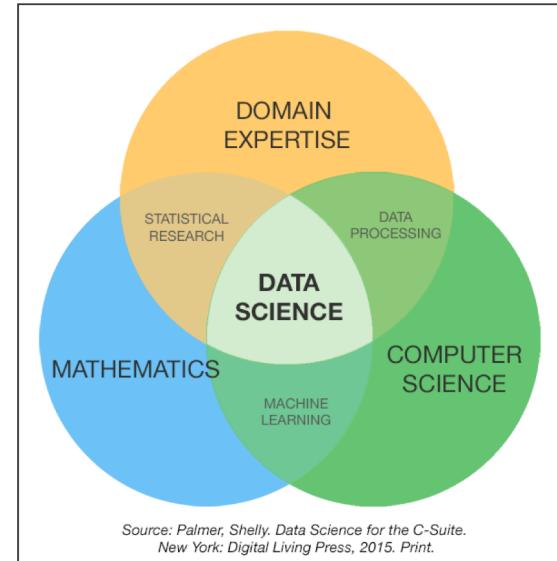
Jake VanderPlas

What is Data Science?

- ☞ **Data science is the discipline of the extraction of knowledge from data.**

It relies on

- computer science
 - for AI, data structures, algorithms, visualization, big data support, and general programming
- statistics/mathematics
 - for data models and inference
- domain expertise
 - for asking questions and interpreting results

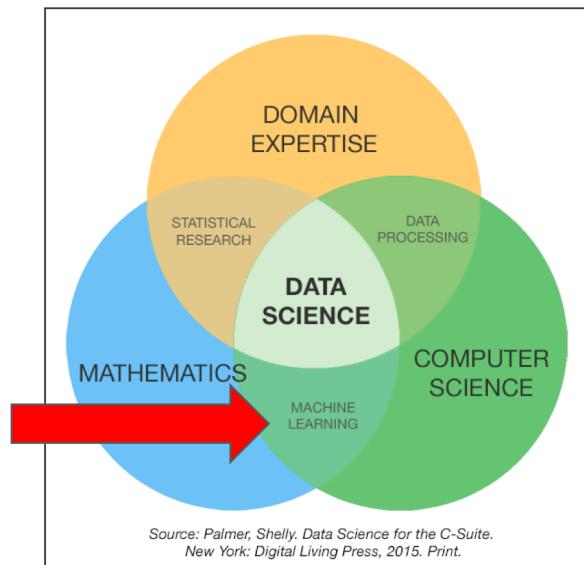


What is Data Science?

- ☞ **Data science is the discipline of the extraction of knowledge from data.**

How do we do that?

- ☞ **We build MODELS of data!**



Models: Play Tennis

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

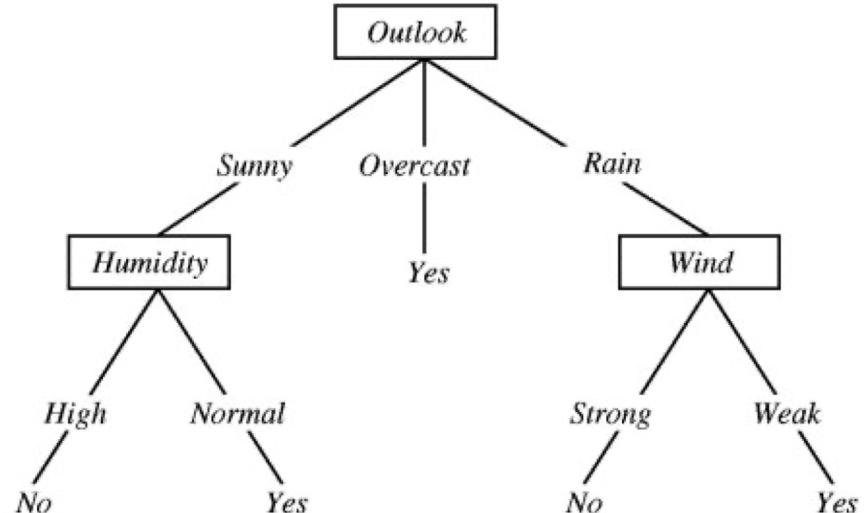
Lots of data - very little information!

Build a model - **a decision tree!**

Models: Play Tennis

ID3 Decision Tree

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No



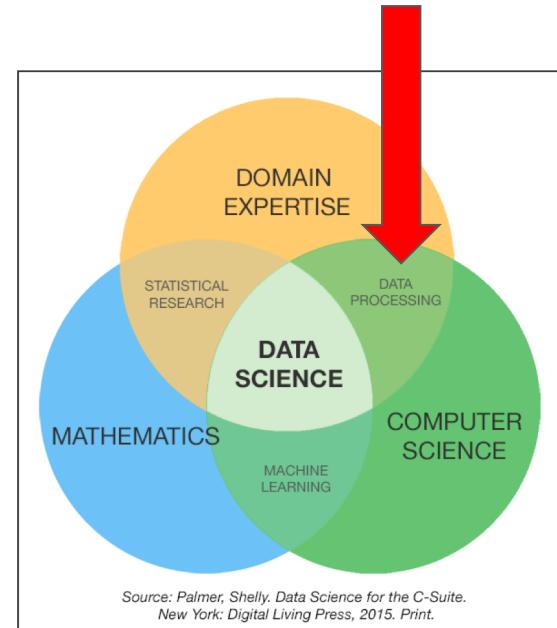
☞ This model summarizes the whole table correctly!

What is Data Science?

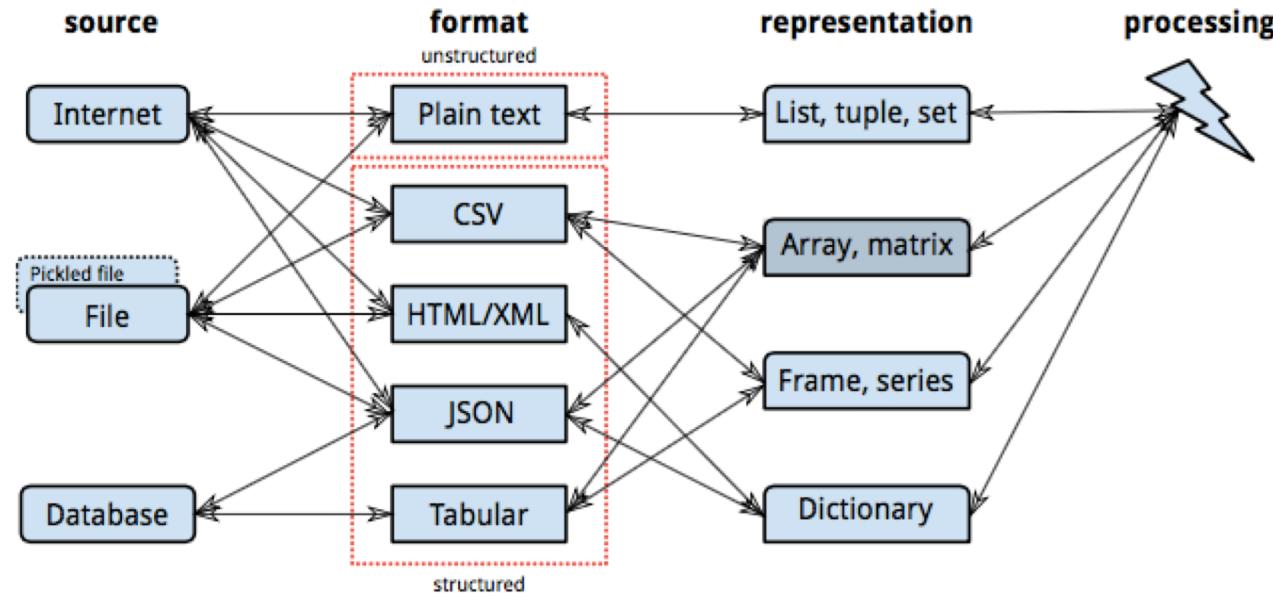
- ☞ **Data science is the discipline of the extraction of knowledge from data.**

Where does the data come from?

- ☞ **The data pipeline!**



The Data Pipeline



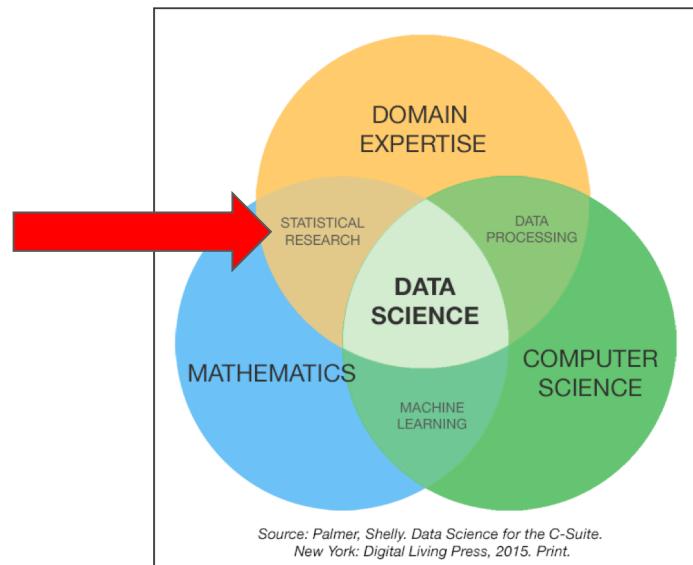
What is Data Science?

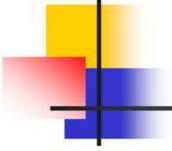
- ☞ Data science is the discipline of the extraction of knowledge from data.

How do we preprocess our data for model building?

- ☞ Statistics!

- Descriptive Statistics
- Missing Value Processing
- Normalization





Descriptive vs. Inferential

- Descriptive Statistics
 - Methods for organizing and summarizing information
- Inferential Statistics
 - Methods for drawing and measuring the reliability of conclusions about a population based on information obtained from a sample of the population

What is Data Science?

- ☞ **Data science is the discipline of the extraction of knowledge from data.**

How do we ask the right questions?

- ☞ **Domain Expertise!**

Knowledge cannot be generated in a vacuum. You need the context of a domain in order to generate new insights. E.g. bioinformatics, climate modeling, sales forecasting, etc.

